Practical Assessment, Research, and Evaluation

Volume 15 *Volume 15, 2010* Article 3

2010

An Overview of Recent Developments in Cognitive Diagnostic Computer Adaptive Assessments

Alan Huebner

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 15, Number 3, January 2010 ISSN 1531-7714

An Overview of Recent Developments in Cognitive Diagnostic Computer Adaptive Assessments

Alan Huebner, ACT, Inc.

Cognitive diagnostic modeling has become an exciting new field of psychometric research. These models aim to diagnose examinees' mastery status of a group of discretely defined skills, or attributes, thereby providing them with detailed information regarding their specific strengths and weaknesses. Combining cognitive diagnosis with computer adaptive assessments has emerged as an important part of this new field. This article aims to provide practitioners and researchers with an introduction to and overview of recent developments in cognitive diagnostic computer adaptive assessments.

Interest in psychometric models referred to as cognitive diagnostic models (CDMs) has been growing rapidly over the past several years, motivated in large part by the call for more formative assessments made by the No Child Left Behind Act of 2001 (No Child Left Behind, 2002). Rather than assigning to examinees a score on a continuous scale representing a broadly defined latent ability as common item response theory (IRT) models do so effectively, CDMs aim to provide examinees with information concerning whether or not they have mastered each of a group of specific, discretely defined skills, or attributes. These skills are often binary, meaning that examinees are scored as masters or non-masters of each skill. For example, the skills required by a test of fraction subtraction may include 1) converting a whole number to a fraction, 2) separating a whole number from a fraction, 3) simplifying before subtracting, and so forth (de la Torre & Douglas, 2004), and a reading test may require the skills 1) remembering details, 2) knowing fact from opinion, 3) speculating from contextual clues, and so on (McGlohen & Chang, 2008). Thus, CDMs may potentially aid teachers to direct students to more individualized remediation and help to focus the self-study of older students.

More formally, CDMs assign to each examinee a vector of binary mastery scores denoted $\boldsymbol{\alpha} = (\alpha_1\ \alpha_2 \ldots \alpha_K)$ for an assessment diagnosing K skills. For example, for K=3, an examinee assigned the vector $\boldsymbol{\alpha} = (1\,0\,1)$ has been deemed a master of the first and third skills and a non-master of the second skill. Since each of the K skills may be assigned two levels, there are 2^K possible skill mastery patterns, which are referred to as latent classes, since mastery and non-mastery are regarded as unobserved categories for each skill. Figure 1 lists all the possible latent classes an examinee may be classified into for K=3 skills, ranging from mastery of none of the skills to mastery of all the skills.

{0 0 0} {1 0 0} {0 1 0} {0 0 1} {1 1 0} {1 0 1} {0 1 1} {1 1 1}
Figure 1: Latent classes for diagnosing K=3 skills

Methods by which examinees are assigned skill mastery patterns will be discussed later in the paper. Some researchers have argued that a binary mastery classification is too restrictive and does not adequately reflect the way students learn; there should be at least one intermediate state between mastery and non-mastery representing some state of partial mastery. While some CDMs are able to accommodate more than two levels of skill mastery, the majority of research has focused on CDMs that diagnose binary skill levels.

While earlier CDM literature focused primarily upon theoretical issues such as model estimation, there

has recently been an increasing amount of work being done on issues that are intended to facilitate practical applications of the models, such as the reliability of attribute-based scoring in CDMs (Geirl, Cui, & Zhou, 2009), automated test assembly for CDMs (Finkelman, Kim, & Roussos, 2009), and strategies for linking two consecutive diagnostic assessments (Xu & von Davier, 2008). In addition, researchers have also been striving to develop the theory necessary to implement cognitive diagnostic computer adaptive assessments, which we refer to as CD-CAT. Jang (2008) describes the possible utility of CD-CAT in a classroom setting with the following scenario. Upon the completion of a unit, a classroom teacher selects various items to be used in a CD-CAT diagnosing specific skills taught in the unit. Students complete the exam using classroom computers, and diagnostic scores are immediately generated detailing the strengths and weaknesses of the students. This vision illustrates the potential of CD-CAT to become a powerful and practical measurement tool. The purpose of this article is to highlight advances in the development of CD-CAT and point out areas that have not been addressed as thoroughly as others. The organization of this article will parallel that of Thompson (2007), who discusses variable-length computerized classification testing according to an outline due to Weiss and Kingsbury (1984), who enumerate the essential components of variable length CAT:

1. Item response model
2. Calibrated item bank
3. Entry level (starting point)
4. Item selection rule
5. Scoring method
6. Termination criterion

It is hoped that some pragmatic information will be provided to practitioners wishing to know more about CD-CAT, and since some of the sections are applicable to CDMs in general rather than only CD-CAT, this article may also serve as a primer to those readers brand-new to the subject.

Psychometric Model

Much of the research into CDMs over the past decade has focused upon the formulation and estimation of new models and families of models. CDMs that have been used in recent CAT research include the Deterministic Input, Noisy-And gate (DINA) model (Junker & Sijtsma, 2001), the Noisy Input, Deterministic-And gate (NIDA) model (Maris, 1999), and the fusion model (Hartz, 2002; Hartz, Roussos, & Stout, 2002). These models vary in terms of complexity, including the number of parameters assigned to each item and the assumptions concerning the manner in which random noise enters the test taking process. In particular, the DINA model has enjoyed much attention in the recent CDM literature, due in large part to its simplicity of estimation and interpretation. It is beyond the scope of this article to provide an in-depth discussion of any specific model; for an overview and comparison of these and various other CDMs see DiBello, Roussos, and Stout (2007) and Rupp and Templin (2008b).

The vast majority of CDMs, including those mentioned above, utilize an item to skills mapping referred to as a Q matrix (K. Tatsuoka, 1985). The Q matrix is an efficient representation of the specific skills that are required by each item in the item bank. For skills k=1… K and an item bank consisting of m=1… M items, the Q matrix entry q_{mk} is defined as

$$q_{mk} = \begin{cases} 1 & \textit{if item m requires skill k} \\ 0 & \textit{otherwise} \end{cases}$$

Thus, each item in the bank contributes exactly one row to the Q matrix. For example, we consider the following Q matrix

$$Q = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

It can be seen that the first item in the bank requires skills 1 and 2, the second item requires skills 1, 3, and 4, the third item requires skill 3 only, and so on. The Q matrix is often constructed by subject matter experts (SMEs), and understandably, much effort has been spent studying this important component of CDMs. For example, Rupp and Templin (2008a) explored the consequences of using an incorrect, or mis-specified Q matrix, de la Torre (2009) developed methods of empirically validating the Q matrix under the DINA model, and de la Torre and Douglas (2008) devised a scheme involving multiple Q matrices for modeling different problem solving strategies.

In addition to determining which skills are required by each item, the SME must also decide how mastery of the skills affects the response probabilities. For example, does a high probability of success result only

DOI: https://doi.org/10.7275/7fdd-6897

when an examinee has mastered all of the required skills or when at least one skill is mastered? Does the probability of a correct response increase gradually as more required skills are mastered? Models demanding that all required skills be mastered for a high probability of a correct response are referred to as conjunctive models; models demanding only some proper subset of the required skills be mastered are called disjunctive. In addition to deciding on a model based upon expert judgment, the response data may be fit to multiple models, and general fit indices such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) may be computed to compare model fit (de la Torre & Douglas, 2008).

In general, there has been no general endorsement of one CDM being better suited for use in CD-CAT applications than any other. Selection of a specific CDM for use in a given assessment will be decided upon by collaboration between SMEs and psychometricians. Clearly, the construction of the Q matrix is of utmost importance for any CDM application, regardless of the specific model used. Finally, in practice a CDM may have to be chosen depending on the computing resources available for estimating the model, which is considered in the next section.

Calibrated Item Bank

Estimating the item parameters of a CDM is generally achieved by an expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977) approach or by Markov Chain Monte Carlo (MCMC) techniques (Tierney, 1994). Examples of models fit by the EM algorithm include the DINA (de la Torre, 2008), the NIDA (Maris, 1999), and the general diagnostic model (GDM) of von Davier (2005), and MCMC has been used to fit models including, but not limited to, the DINA and NIDA (de la Torre & Douglas, 2008) and the fusion model (Hartz, 2002). These papers outline algorithms which may be implemented by practitioners in the programming language of their choice, or existing ready-made software packages may be utilized. Such programs include Arpeggio (Educational Testing Service, 2004), a commercial package which estimates the fusion model and a routine for use in the commercial software M-Plus (Muthén & Muthén, 1998-2006) which estimates a family of CDMs based upon log linear models (Henson, Templin, & Willse, 2009). A list of various commercial and freeware software programs for estimating CDMs may be found in Rupp and Templin (2008b).

There are some complications, however. Not all of the software is well documented, and some programs are available only to researchers. An issue critical to the practical implementation of an operational CD-CAT program is that the algorithms described in the above papers and some of the software is designed for full response matrices only and must be modified by the practitioner to handle response data in which items are not seen by every examinee. Another practical concern is computing time; in general, the EM algorithm will converge much more quickly (especially when diagnosing a small number of skills) than MCMC methods, for which convergence may take several hours or even possibly days. For this reason, as well as the extreme care required to assess the convergence of the parameters estimated via a MCMC algorithm, practitioners may conclude that the EM algorithm approach is the preferable estimation method in the context of an operational diagnostic assessment program.

There have been few concrete recommendations in the literature regarding minimum sample size for calibrating item parameters for CDMs. Rupp and Templin (2008b) suggest that for simple models such as the DINA a few hundred respondents per item is sufficient for convergence, especially if the number of skills being diagnosed is not too large, such as four to six. A systematic study investigating minimum sample size for item calibration for different CDMs and for various numbers of skills is currently lacking. A related issue is that of model identifiability, or the property of the model that ensures a unique set of item parameters will be estimated for a given set of data. von Davier (2005) states that models diagnosing greater than eight skills are likely to have problems with identifiability, unless there are a large number of skills measuring each item. For a simple example of how such problems might arise, consider attempting to estimate a model diagnosing K=10 skills using a sample of N=1000 examinees. Since the number of possible latent classes (2^{10}=1024) is greater than the actual number of examinees, it is doubtful that accurate parameter estimates and examinee classifications will be obtained. Of course, models having fewer parameters per item will have less difficulty with identifiability than models with more complex parameterizations, and again, there have been no systematic studies for CDMs investigating the relationships between identifiability, sample size, and the number of skills being diagnosed.

Starting Point

The issue of the selection of items that are initially administered to examinees at the start of a CD-CAT assessment has not been explicitly addressed. In their simulation study Xu, Chang, & Douglas (2003) begin the simulated exams by administering the same set of five randomly chosen items to each examinee. If examinees are subjected to a series of diagnostic exams, such as a pretest/test/retest scheme, then it would be possible to start the exam by selecting items (see the next section) according to the examinee's previous classification. Whether selecting initial items in this fashion or randomly affects the examinee's ultimate classification is currently unknown.

Item Selection Rule

Much of the CDM literature that is specific to CD-CAT applications focuses upon rules for item selection. Several rules and variations have been proposed for both assessments that are designed to exclusively provide diagnostic information and for assessments that provide an IRT theta estimate as well as diagnostic results. Concerning the former scenario, Xu et al. (2003) apply the theoretical results of C. Tatsuoka (2003) to a large scale CD-CAT assessment using the fusion model. Two item selection procedures are proposed; a procedure based upon choosing the item from the bank which maximizes the Kullback-Leibler (KL) information, a measure of the distance between two probability distributions, and a procedure based upon minimizing the Shannon Entropy (SHE), a measure of the flatness of the posterior distribution of the latent classes (see the next section). It is shown that, for fixed length exams, selecting items via the KL information or SHE leads to higher classification accuracy rates compared to selecting items randomly. The SHE procedure is slightly more accurate than the KL information, but with more skewed item exposure rates. Cheng (2009) proposed two modifications to the KL information procedure, the posterior weighted Kullback-Leibler (PWKL) procedure and the hybrid Kullback-Leibler (HKL) procedure. Both were shown to yield superior classification accuracy compared to the standard KL information and SHE procedures. One note of practical concern is the computational efficiency of these various item selection rules. The KL information procedure is by far the most efficient, since information has to be computed only once for a given item bank. On the other hand, the SHE procedure requires that considerable calculations be performed over every remaining item in the bank each time an item is administered.

Item selection procedures have also been proposed for the case in which both a common IRT model and a CDM are fit to the same data in an attempt to simultaneously estimate a theta score and glean diagnostic information from the same assessment. McGlohen and Chang (2008) fit the three parameter logistic (3PL) and the fusion models to data from a large scale assessment and simulated a CAT scenario in which three item selection procedures were testing. The first procedure selected items based upon the current theta estimate (via maximizing the Fisher information) and classified examinees at the end of the exam, the second procedure selected items based upon the diagnostics (via maximizing the KL information) and estimated theta at the end, and the third procedure selected items according to both criterion by the use of combining shadow testing, a method of constrained adaptive testing proposed by van der Linden (2000), and KL information. The first and third procedures displayed good performance for both the recovery of theta scores and diagnostic classification accuracy.

Scoring Method

Examinee scoring in the context of CDMs involves classifying examinees into latent classes by either maximum likelihood or maximum posteriori. There is no distinction between obtaining an interim classification during a CD-CAT and a classification at the end of a fixed length diagnostic exam. We will demonstrate the maximum posteriori method, since the maximum likelihood method is equivalent to a special case of maximum posteriori. For an assessment diagnosing K skills, the i^{th} examinee is classified into one of the 2^K possible latent classes given his or her responses, denoted X_i, and the set of parameters corresponding to the items to which the examinee was exposed, denoted ϕ_i. The likelihood of the responses given membership in the l^{th} latent class and the item parameters may be denoted as $P(X_i \mid \alpha_l, \phi_i)$, and the prior probability of the l^{th} latent class is denoted as $P(\alpha_l)$, which may be estimated from a previous calibration or expert opinion. Then, the desired posterior probability $P(\alpha_l \mid X_i)$, the probability of the i^{th} examinee's membership in the l^{th} latent class given

her response sequence, may be found using the formula (Bayes Rule)

$$P(\alpha_l \mid X_i) = \frac{P(X_i \mid \alpha_l)P(\alpha_l)}{\sum_{c=1}^{L} P(X_i \mid \alpha_c)P(\alpha_c)}.$$

Calculating the posterior distribution of the latent classes entails simply using the above formula for all l=1... L possible latent classes. The examinee is then classified into the latent class with the highest posterior probability. When the value $1/L$ is substituted for $P(\alpha_l)$ in the computation, referred to as a flat or non-informative prior, the result is equivalent to classification via maximum likelihood.

Upon the completion of a CD-CAT assessment, it may be desired to provide the examinee with a graph of individual skill probabilities, or skill "intensities," in addition to simple binary mastery/non-mastery classifications. Such a graph may be constructed using the final posterior distribution of the latent classes. For example, suppose a hypothetical examinee is administered a CD-CAT assessment diagnosing K=3 skills and upon completion of the exam the posterior distribution shown in Table 1 is computed based upon the responses and item parameters of the exposed items. Clearly, the examinee would be assigned the mastery vector {1 0 1}, since this class has the highest value in the posterior distribution.

However, we may also compute the probability that the examinee has mastered each individual skill. Since the latent classes are mutually exclusive and exhaustive, we may simply add the probabilities of the latent classes associated with each skill. Specifically, denote the probability that an examinee has mastered skill k as *P(skill k)* and the probability that the examinee is a member of latent class $\{\alpha_1 \alpha_2 \alpha_3\}$ as $P(\{\alpha_1 \alpha_2 \alpha_3\})$. Then

$$\begin{aligned} P(skill\ 1) &= P(\{1\ 0\ 0\}) + P(\{1\ 1\ 0\}) + P(\{1\ 0\ 1\}) + P(\{1\ 1\ 1\}) \\ &= 0.15 + 0.05 + 0.43 + 0.13 \\ &= 0.76 \end{aligned}$$

Similar calculations yield *P(skill 2)*=0.24 and *P(skill 3)*=0.72. These probabilities may be expressed via a bar graph as in Figure 2. These graphs may help students and teachers grasp diagnostic results in a more intuitive fashion than classification alone.

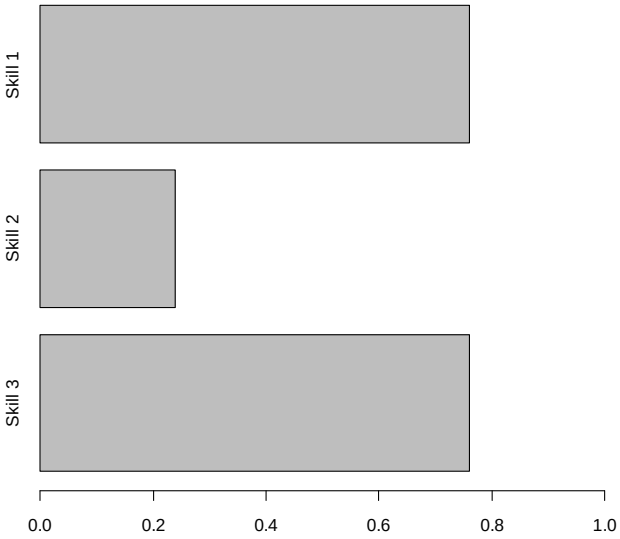

Figure 2: Graph of individual skill probabilities

Termination Criterion

In general, discussions of termination criteria, or stopping rules, for CD-CAT have been largely absent from the current literature. One exception is C. Tatsuoka (2002). Working in the context of diagnostic classification using partially ordered sets, an approach in which examinees are classified into "states" rather than latent classes and thus somewhat different than that taken by the CDMs discussed in this paper, he proposes that a diagnostic assessment be terminated when the posterior probability that an examinee belongs to a given state exceeds 0.80.

Table 1: Posterior probability for hypothetical examinee

Latent class	{0 0 0}	{1 0 0}	{0 1 0}	{0 0 1}	{1 1 0}	{1 0 1}	{0 1 1}	{1 1 1}
Posterior probability	0.06	0.15	0.02	0.12	0.05	0.43	0.04	0.13

This concept may be easily adapted to CDMs by terminating the exam when the probability an examinee belongs to a latent class exceeds 0.80, and this threshold may be lowered or raised if it is desired to sacrifice some classification accuracy in exchange for shorter exams, or vice versa. This stopping rule, and likely other stopping rules for CD-CAT yet to be proposed, utilizes the posterior distribution of the latent classes as a measure of the precision of classification, similar to the standard error on an IRT theta estimate. The more "peaked" a distribution is at one class, the more reliable the classification will be. Clearly, a termination rule which stops a CD-CAT exam when an examinee is assigned posterior distribution in Table 2 will most likely yield more accurate classifications than a rule which stops the exam when the posterior distribution is similar to that shown in Table 1 for the previous example. The performance of Tatsuoka's termination rule at thresholds higher and lower than 0.80 in terms of classification accuracy and test efficiency, as well as the formulation of new termination rules, may prove to be fruitful directions for research.

Discussion

CDMs are statistically sophisticated measurement tools that hold great promise for enhancing the quality of diagnostic feedback provided to all levels of students in many different types of assessment situations. New models, both simple and complex, that measure various cognitive processes are rapidly being proposed, and means of estimating these models are being made more and more accessible to practitioners. In order for CDMs to fulfill their potential, however, researchers must still answer basic general questions regarding concerns such as the reliability and validity of the results yielded by CDMs. For example, for simulation studies in which response data are generated to fit a given model exactly, CDMs are capable of classifying individual skill masteries with over 90% accuracy (de la Torre & Douglas, 2004; von Davier, 2005). However, there is less understanding as to how accurately examinees are classified in real world applications, i.e., when the examinee responses do not fit a given model exactly.

Questions also remain that are specific to CD-CAT. In order for Jang's (2008) hypothetical scenario detailed above to become a reality, CD-CAT assessments must be made to be efficient, accurate, and sufficiently uncomplicated so that they may be effortlessly incorporated into actual classrooms. This article has aimed to describe areas of CD-CAT methodology that are being developed to a high degree, such as item selection rules, as well as areas which remain somewhat unexplored, such as termination rules. It is hoped that some useful direction has been provided to practitioners wishing to begin working and experimenting with this new methodology.

References

Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika.* Advance online publication. doi: 10.1007/s11336-009-9125-0

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, *34*, 115-130.

de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika. 69*(3), 333-353.

de la Torre, J., & Douglas, J. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*,*73*(4), 595-624.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 39*, 1, 1-38.

DiBello, L., Roussos, L., & Stout, W. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C.R Rao & S. Sinharay (Eds.) *Handbook of Statistics, 26*, (pp. 979-1030). Amsterdam: Elsevier.

Educational Testing Service (2004). Arpeggio: Release 1.1 [Computer software]. Princeton, NJ: Author.

Finkelman, M., Kim, W., & Roussos, L. (2009). Automated test assembly for cognitive diagnostic models using a genetic algorithm. *Journal of Educational Measurement, 46 (3)*, 273-292.

Gierl, M., Cui, Y., & Zhou, J. (2009). Reliability and attribute-based scoring in cognitive diagnostic

Table 2: Example of a "peaked" posterior distribution.

Latent class	{0 0 0}	{1 0 0}	{0 1 0}	{0 0 1}	{1 1 0}	{1 0 1}	{0 1 1}	{1 1 1}
Posterior probability	0.00	0.02	0.01	0.02	0.06	0.85	0.03	0.01

assessment. *Journal of Educational Measurement, 46 (3)*, 293-313.

Hartz, S. (2002). *A Bayesian framework for the Unified Model for assessing cognitive abilities: blending theory with practice.* Unpublished doctoral thesis, University of Illinois at Urbana-Champain.

Hartz, S., Roussos, L., & Stout, W. (2002). *Skills diagnosis: Theory and practice* [User manual for Arpeggio software]. Princeton, NJ: Educational Testing Service.

Henson, R., Templin J., & Willse J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika, 74*(2), 191-210.

Jang, E. (2008). A framework for cognitive diagnostic assessment. In C.A. Chapelle, Y.-R. Chung, & J. Xu (Eds.), *Towards an adaptive CALL: Natural language Processing for diagnostic language assessment* (pp.117-131). Ames, IA: Iowa State University.

Junker, B., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258-272.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*(2), 187-212.

Muthén, L.K., & Muthén, B.O. (1998-2006). *M-plus user's guide* (4th ed.). Los Angeles: Muthén, L.K., & Muthén.

McGlohen, M., & Chang, H. (2008). Combining computer adaptive testing technology with cognitively diagnostic assessment. *Behavior Research Methods, 40 (3)*, 808-21.

No Child Left Behind Act of 2001, Pub. L. No. 107-110 Stat. 1425 (2002).

Rupp, A., & Templin, J. (2008a). The effects of *q*-matrix misspecification on parameter Estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement, 68(1)*, 78-96.

Rupp, A., & Templin, J. (2008b). Unique characteristics of diagnostic classification models: a comprehensive review of the current state-of-the-art. *Measurement,6*, 219-262.

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Applied Statistics, 51*(3), 337-350.

Tatsuoka, C., & Ferguson, T. (2003). Sequential classification on partially ordered sets. *Journal of the Royal Statistical Society, Series B, 65*(1), 143-157.

Tatsuoka, K. (1985). A Probabilistic Model for Diagnosing Misconceptions in the Pattern Classification Approach. *Journal of Educational Statistics, 12*, 55-73.

Thompson, N. (2007). A practitioner's guide for variable-length computerized classification testing. *Practical Assessment, Research & Evaluation*, *12*(1), 1-13.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics, 22,* 1701-1786.

von Davier, M. (2005). *A General diagnostic model applied to language testing data*. ETS Research Report. Princeton, New Jersey: ETS.

van der Linden, W. (2000). Constrained adaptive testing with shadow tests. In W.J. van der Linden & C.W. Glas (Eds.) *Computerized adaptive testing: Theory and practice* (pp. 27-52). The Netherlands: Kluwer Academic Publishers.

Weiss, D., & Kingsbury, G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, *21*(4), 361-374.

Xu, X., Chang, H., & Douglas, J. (2003). *A simulation study to compare CAT strategies for cognitive diagnosis.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Xu, X. & von Davier, M. (2008). *Linking for the general diagnostic model.* ETS Research Report. Princeton, New Jersey: ETS.

Citation

Huebner, Alan, (2010). An Overview of Recent Developments in Cognitive Diagnostic Computer Adaptive Assessments. *Practical Assessment, Research & Evaluation*, 15(3). Available online: http://pareonline.net/getvn.asp?v=15&n=3.

Author

Alan Huebner
ACT, Inc.
500 ACT Drive, P.O. Box 168
Tel: 319-341-2296
Fax: 319-337-1665
alan.huebner [at] act.org